GLOBAL INFO - the German Digital Library Project: development and perspectives

Schmiede, Rudi

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Rudi Schmiede March 1998

GLOBAL INFO – The German Digital Library Project

Development and Perspectives

Electronic information and communication in the sciences have spread rapidly in recent years. Electronic catalogues of books and other documents, full texts, or databases of facts are becoming accessible on a national and international scale. World-wide retrieval and communication, electronic publishing and cooperation are widely used in natural and technological sciences, but also in medicine, whereas in humanities and social science they have only recently started to become a common medium.

This is the reason why in Germany in 1995 a cooperation agreement between some important professional scientific societies –the mathematicians, the physicists, the chemists, and the information scientists – was agreed upon and resulted in the so-called "common initiative of German scientific societies for electronic information and communication" (IuK-Initiative). The yearly conference of this initiative was recently – in March 1998 – convened in Hamburg and presented an ambitious international programme on "Integrated Scientific Information Systems". In 1996 and 1997, other scientific societies – that of pedagogics, of sociology, of biology, of electrical engineering science, and of psychology – have joined this agreement so that a considerable part of German science is taking part in the initiative. This is the background from which – together with paralleling activites of publishing houses and others – the idea of the funding programme GLOBAL INFO was developed in 1997.

This programme, which is to run for six years (1998 – 2003) and has a planned financial volume of about 60 Mill. DM, includes all the disciplines whose learned societies are active in the IUK initiative. The most important aims of the IuK initiative have been integrated into GLOBAL INFO. The programme shall – as it is said in the ministry's press release –

- advance “optimal access to the world-wide electronic and multimedial information on full texts, literature references, factual databases and software” for the single scientist.

- This access to bases of information “which are stored in distributed information systems, shall be accessible from their work-place computer. Keyword is the 'digital library‘.”

- The programme aims at the cooperation of all parties acting in the process of providing, distributing and using information; it is to integrate the producers (i.e. the authors, represented by the learned societies or the scientific publishing houses), the distributors (publishing companies, book sellers, scientific database providers, and scientific libraries) and the consumers (i.e. readers or users, represented again by the learned societies and representatives of university departments).

- Cooperation of these parties in project consortia in the precompetitive sphere is the condition of funding for projects.

If one considers the technical, economic, social and political frame within which GLOBAL INFO is gaining shape it seems to be very clear that this programme will either adjust to the international standards and development in distributed information infrastructure (i.e. have real global dimensions) or it will perish. The basic global medium of the existing distributed information infrastructure is the internet and especially the WWW. GLOBAL INFO will have to be compatible to this structure, i.e. to develop elements and single systems being connective to the Web [Fig. 1]. Within these modular single systems the dimensions of protocols (e.g. HTML, Z.39.50 etc.), of metadata (e.g. thesauri, classifications of various kinds etc.), and of heterogeneity (relevance, consistence, scope) are not independent of each other, but are one system structure. These systems should be simply structured, lean and flat (with few abstraction layers). Because of the interrelated character of the dimensions of this modular systems, they may be proprietary as long as the condition of connectivity is fulfilled. Within these systems distributed objects should support a multidimensional functionality [Fig. 2]. Since the distributed system at large is the Web there will not be *the* system architecture of GLOBAL INFO. A minimum number of protocols, interfaces, and requirements for interoperability has to be agreed upon (but they, too, are essentially determined by global standards).

The aim of GLOBAL INFO, therefore, can be defined in the following way: The programme is to advance – in the context of a global, principally open system of information provision and brokering – the development and implementation of modular user-oriented single systems for various functionalities.

In order to make sure that these principles will be realized by the many activities which have emerged rapidly the GLOBAL INFO CONSORTIUM was installed as board of GLOBAL INFO to guide the programme as a whole and to decide upon the essentials of its direction and organization.

The GLOBAL INFO CONSORTIUM (GIC) consists of 5 representatives of learned societies, 4 representatives of publishing houses, 1 representative of scientific information centers and 1 representative of scientific libraries. Representatives of the Ministry (BMBF), the programme organizing institution (PTF) and the German research funding society (DFG) have guest status at the GIC [Fig. 3]. The secretary of GLOBAL INFO resides with the Society of German Chemists (GDCh) at Frankfurt; the GLOBAL INFO server is located physically and administered technically at the university of Oldenburg [*http://www.global-info.org/*]. Responsible for this organization is the chairman, supported by an assistant who cooperates closely with the secretary and the server administrator.

The GLOBAL INFO CONSORTIUM has confirmed the conditions of

- cooperation of all parties and of

- various scientific disciplines
  stated by the ministry. It has stressed the

- limitation of development activities to the pre-competitive area; this excludes singular development projects which are specific for one scientific discipline, for one location or one company.
  Furthermore, it has emphasized the

- crucial importance of the international orientation of work in order to avoid national cul-de-sacs, and of the necessity to evaluate every single activity with respect to the international “state-of-the-art”, i.e. the world market standard.
  To accomplish that, it has stated that every project has to undergo a procedure of

- external and – as far as possible – international evaluation of its perspective.

To solicitate the move towards a network integrating the various parties, scientific disciplines and local activities the GLOBAL INFO CONSORTIUM has installed 5 working parties [Schwerpunkte] which – in the period of transition to a general network – are to organize in a pragmatic way the development of a cooperative structure of all partners. Each of them is coordinated by two GLOBAL INFO CONSORTIUM members, one from science and the other from providers or distributors; they are supported by steering commitees. Ideally, this work should result in the transgression of these working parties by a network of cooperative projects – organized in consortia – across the borderlines of parties and disciplines. This move is supported by so-called pre-projects [Vorprojekte] which offer a maximum of 100.000 DM over 2 years to invest into the building of cooperative structures in the mentioned sense. 16 of these projects have been granted to universities, 13 to publishing houses for the period 1998/99. Above that, several institutions with special interests in this field (mainly foreign publishing companies, German science database providers and scientific libraries) have been invited to join the programme. In the first wave of project development these groups of participants (universities and publishing companies with pre-projects and acknowledged interest institutions) are intitled to hand in project proposals (under the conditions stated above).

GLOBAL INFO is in the middle of this process towards a network of cooperation now. The 5 working parties have met for several workshops. Usually, representatives of the other working parties were present at each workshop so that first strings of cooperation have been installed. The GLOBAL INFO CONSORTIUM recently decided that a first deadline for project proposals should be by July 31st, 1998 (this is exactly 1 year after the deadline for proposals of pre-projects). Viewing the current status of work in the working parties this date should be realistic to receive the first series of really integrated propositions.

The contemporary status of work in the 5 working areas can be summarized in the following way:

Working Party I: Completion and processing of content: document types, process and tools for electronic publishing, information transfer, storage, conversion and indexing.

In the last meeting of this working party more than 20 different single projects have been clustered into 4 resp. 5 focusses with several cooperators planning to take part in this activity:

1. **Formats + Markup** (Brüggemann-Klein of TU München Verlag, Verlag Leske & Budrich, Univ. Tübingen, dPunkt-Verlag, Spektrum Akademischer Verlag)

2. **Formats + Workflow** (Ziegler of FIZ Karlsruhe, Univ. Bonn, Westdeutscher Verlag)

3. **Multimedia-Authoring** (Hogrefe Verlag, FH Regensburg, Teubner-Verlag, Cornelsen Software, OFFIS Oldenburg, Univ. Freiburg)

4. **"Chemistry" + dynamic documents** (Univ. Oldenburg, Univ. Tübingen, CASAF Bitterfeld)

5. **Archiving (still open whether separate focus)** (Die Deutsche Bibliothek, SUB Göttingen, Bayerische Staatsbibliothek)

It is yet to decide whether the problem of archiving is to become a separate focus, or whether it will be a string parallel to the other 4 focusses (as it is with the print model advanced by HWTK Leipzig). It is clear, however, that archiving parallels the whole production process not only the product phase.

Working Party II: Networking of educational materials

In this area of work, three focusses have been defined:

1. Learning environments, teaching and learning material

2. Multimedia tools for authors, exemplary applications

3. Metadata, retrieval, integration of information and communication

Project proposals have been discussed for the following areas:

- **Tools** (SpringerVerlag, Verlag Harry Deutsch, Univ. Osnabrück, Univ. Tübingen etc.)

- **Exemplary learning scenarios** (Univ. Oldenburg, Univ. Dresden, Univ. Osnabrück)

- **Learning with electronic media** (participants not yet clear)

- **Metadata and metadata server, document server of learning materials, services** (HU Berlin, Spektrum Akad. Verlag, Univ. Jena, Univ. Kaiserslautern, Univ. Osnabrück, TU Dresden, FH Heilbronn, Cornelsen Software, DIPF Frankfurt/Main)

Working Party III: Formal description, identification and retrieval, metadata, networking

This working area aims at the

- formal and content-oriented description of digital objects,
- identification of resources and their retrieval by intelligent search engines,
- networking of digital objects and
- terms and conditions.

To make these general aims operable, 3 areas of work have been defined:

1. **Metadata** (enhancement of consistency and quality)

2. **Dealing with heterogeneity**

3. **Enhancement of functions**

Concrete proposals for projects have been developed for the subjects of

- Integration of data sources

- Varying search, retrieval and information systems
- Diverse forms to develop and to deal with queries
- Gathering and brokering
- Metadata repositories
- Administration of Uniform Resource Names (URNs)

In addition a list of items and tasks which should become part of the party's work has been put together: Metadata tags for various documents; interoperable metadata for the description of sources; tools to transform formats carrying metadata (HTML/RDF/XML); cooperation in international norming and conversion of international developments into the projects to avoid national peculiarities; search machines for heterogeneous sources; gatherer tools; metadata repositories with common exchange formats; transparent offers of research processes to users; essencing tools, automatic classification, thesauri; presentation, reorganization, and annotation of research results.

In this working party, however, the projects – let alone the ideas – cannot yet be attributed to authors resp. research teams.

<u>Working party IV</u>: Usage of content: alerting, awareness, information pooling, information mediation, evaluation of results, platforms, intelligent agents, access (passwords)

This working area is the most complex and, therefore, the most difficult to structure. In a way, it is the central area of GLOBAL INFO, since it contains key elements of the system architecture in the sense of conventions about protocols, interfaces, i.e. interoperability. There is a broad consense that a *common open distributed system* has to be defined and modular parts of it have to be developed; it may use the experience and, possibly, integrate parts of the former Medoc project of information science (which is to be integrated into GLOBAL INFO by a transition project called *Interdoc*). The perspective of developing rapidly prototypes for certain services by using existing testbeds has gained some adherence in this working area.

In this working party many persons, institutions, and projects are present so that meetings of this working party resemble large workshops. This is the reason why in its last meeting three weeks ago at Heidelberg this working party decided to meet not longer in toto but to work in parallel groups coordinated by a steering commitee. The project proposals are just being collec-

ted resp. re-formulated. They might be grouped in two large groups – one related to services to be realized, the other one adressed to more general problems of architecture – as follows:

1. Services

- **Research and Retrieval** (Univ. Bochum, TU Darmstadt, IZ Bonn, Springer Verlag, Fuhr of Univ. Dortmund, Günthner of Univ. München – close relation to WP III)
- **Indexing and Content Analysis** (Univ. Osnabrück, Lehmanns Buchhandlung, Ley of Univ. Trier, Helbig of FU Hagen – close relation to WP III)
- **Alerting and Profile Services, push and pull techniques, personal virtual library** (FU Berlin, Univ. Stuttgart, Univ. Tübingen, Zimmermann of Univ. Saarbrücken)
- **Global Password** (TU Darmstadt)
- **Legacy Services** (Interdoc)
- **Visualization, Browsing and Navigation** (Univ. Stuttgart, TU Darmstadt, Univ. Tübingen, Mie of GFZ Potsdam)
- **Interoperability Platform** (Univ. Hamburg, TU Darmstadt, FU Berlin)

2. General Problems

- **User Questionnaire** (Univ. Hamburg, TU Chemnitz, Elsevier Publishers)
- **User desktop** (TU Chemnitz)
- **Integration of E-Journals** (FU Berlin, Univ. Regensburg)
- **Metadata, Classification and Federation Services** (Univ. Oldenburg, UB Regensburg, Univ. Rostock, Saake of FH Magdeburg – close relation to WP III)

Working Party V: Business Models, billing (micro-billing) and payment, statistics

Three working groups have been installed in this working area of 4 – 5 members each:

1. **Units** (speaker: Mr. Marschall, DITR)

2. **Interfaces and functions** (speaker: Mrs. Hotzel, Univ. Jena)

3. **Billing systems** (Mr. Schwab, Springer Verlag)

Up to now, a first round of discussion of project proposals – including the parallel DFN-project CHABLIS (Charging, Accounting and Billing for Digital Library Service) – has taken place. This working party is planning a joint project in cooperation with some of the other focusses.

Recapitulating this overview of the 5 working parties and the various activites and project proposals within them one can see, that GLOBAL INFO is in the process of constructing a rather complex network of interdisciplinary and inter-regional development projects in which the relevant players in this field cooperate. The formal barriers to get these projects funded by the Ministry of Science are substantial: At least 2 universities, 2 publishing houses and in addition other information providers (i.e. database providers, libraries, booksellers etc.) have to form a consortium which is in corporate responsibility to realize the project's aim. The project proposals will have to include a detailled specification of working steps, milestones, and functionalities which have to be realized in the course of development of a module or a larger system.

Probably most of the working parties will need one more meeting among themselves to sort proposals together or out and to intensify contact with related projects in other working parties. At the same time a small group of system specialists has been set up to meet in a kind of "conclave" to define a first raw set of conventions the projects have to respect (i.e. the elements of a system architecture or the essentials of interoperability). These activities will take place in April and May. With the handing in of the first wave of project proposals of consortia by 31st July, 1998, a screening process with several steps is already under way: The proposals of consortia are result of an intensive discussion in and between the working parties, and they a product of a coordination process by the working party coordinators with their steering commites. Then, the GLOBAL INFO CONSORTIUM will discuss and evaluate these proposals. It will propose a number of independent and international specialists to the Ministry out of which the Ministry will select the ones it deems to be apt and convenient. After evaluation by these specialists, the GLOBAL INFO CONSORTIUM will decide whether to recommend or not to recommend the single proposal to be funded. At the end the Ministry will ultimately decide upon the funding.

So after all it seems that GLOBAL INFO by the end of this year will have accomplished the transition into a nation-wide networking system of interdependent projects with intensive international contacts and cooperation towards a distributed and dynamically developing digital library. One can realistically hope that within a few years the chances of the single scientist to use the advantages of large integrated scientific information systems will be amended dramatically. It may be hoped, too, that a new "division of labour" between the traditional players in this field – the scientists themselves, their universities, the libraries, the database providers, the publishing companies and the book sellers – will be achieved; one should not underestimate this economic and political dimension as one of the central problems and tasks in the construction of large information systems. The solution of organizational questions and problems is equally important – it might be even more urgent – as the technological development in this sphere. But all the participants in this process should continuously have in mind that they can only succeed together: The *scientists* will not be able (and should not be interested) to compete the publishing houses out of the market; and the other way around the great *publishers* will have to avoid to dry out their customers' basis. The *database providers* will have to find a rapid and smooth way to integrate their offer into an increasingly integrated – Web-mediated – information landscape; proprietary retrieval forms and languages will have to be substituted by visualized, intuitive forms of research to be handled without special retrieval knowledge. The *libraries* and the *librarians* will have to accomodate with their changing role as local or regional information broking institutions instead of mere book storage institutions; from a self-understanding to be primarily archiving units they will have to move to that of specialists for the provision and brokering of information. The same is – mutatis mutandis – true for the *booksellers* and *bookstores* in the market sphere.

There is yet another respect in which questions of content seem to be rather more complicated than purely technical ones: With the integration of scientific information systems the problem of identifying properly the adequate scientific culture, of developing intelligent description and retrieval systems which will guide me reliably into these areas of science which I am really looking for, will gain momentum. The relation of intellectual input and structuring on the one hand, the possible automation of processes of information providing and retrieval on the other hand are far from being solved. So as another big task we will have to be careful to accomplish a scientific information reality which will allow that in every discipline of science teachers, researchers and students are helped, not restricted by or overflood with huge piles of senseless information by the systems we develop. What a German in-

formation scientist some ten years ago termed the ‚explosion of nonsense' has to be avoided. So questions of conceptual analysis, of the different meanings of terms, i.e. of the special knowledge frameworks and concepts in the differing scientific cultures will have to play their due role in the development of integrated scientific information systems.

With all these questions in view GLOBAL INFO will have to proceed in a tentative pragmatic way. To find a due balance between systematic thinking and conceptualization and the development of practically relevant and rapidly usable solutions and results, is one of the most interesting, but also one of the most difficult aims in science; the same is true for the central task of GLOBAL INFO: the development of useful integrated scientific information systems.